An Integrated Approach towards Building a Simulation Model Supporting the Management of the Passenger Transportation System. Part 1 - Theoretical basis.

Joanna Gąbka, Sławomir Susz, Maria Rosienkiewicz  
*Wrocław University of Technology, Poland*

This article presents an integrated approach towards building a simulation model of the transportation network. The proposed method is based on the recommendations of the Blue Book for Sector of Public Transport in cities, agglomerations and regions issued by Jaspers (Joint Assistance to Support Projects in European Regions). It comprises six steps with such elements as econometric modelling, artificial neural networks and mathematical model. It is dedicated to developing simulation models for the purpose of picturing present situation and dependencies dominating in the transportation network, easily reflecting effects of potential modifications and providing forecasts for the future. The described hybrid method was verified in practice by building the simulation model which is presented in the next article in this series entitled: An Integrated Approach towards Building a Simulation Model Supporting the Management of the Passenger Transportation System. Part 2 – Case Study.  
**Keywords:** integrated approach, management system, transport network, algorithm components, simulation model.

1. INTRODUCTION

Building a reliable simulation model of the Passenger Transportation system that will serve as a useful decision supporting tool is a very complex task. It requires integration of several approaches and proper conclusions from observations and data collected in different areas. Using the correct method of the model constructing decides whether the developed tool is easy to use and exploit by decision makers and authorities responsible for the transportation system management. Because there is an evident need for the decision supporting tools in such a dynamic and complex field as a transportation network, for the purpose of their smooth expansion and modification, an integrated method of model building was developed. The approach presented in this article proved to be reliable, and thus enabled building an effective simulation model for the transportation network. It complies with the best practices presented in the literature as well as with the recommendations contained in the Blue Book for Sector of Public Transport in cities, agglomerations and regions issued by Jaspers (Joint Assistance to Support Projects in European Regions) which is the additional advantage encouraging to its broad use.

2. MODELS AND MODELLING

A model is defined as a simplified representation of a system at some particular point in time or space, intended to promote understanding of the real system. Simulation is the manipulation of a model in such a way that it operates on time or space to compress it, thus enabling one to perceive the interactions that would not otherwise be apparent because of their separation in time or space, [37]. A simulation model is usually characterised by a very complex structure. Crucial elements of this structure are procedures or rules that control the internal mechanisms of the model. These procedures or rules can be equations, mathematical, statistical or econometric models, artificial neural networks and others. Another critical aspect when it comes to the model’s correctness is the proper and exhaustive dataset. Only implementing adequate procedures and a complete set of data in the model assures that the model works appropriately and correctly reflects a simulated phenomenon or a process.
Development of computing possibilities resulted from intensification of research focused on complex system behaviour modelling, including transportation networks and passenger flows. In the literature, the simulation models were categorised into three main categories according to scale. The first group is represented by large-scale macro models which describe traffic flow generally as populations of objects. On the small side of the scale there are micro-models with individual objects defined in detail with their characteristics. In between these two very large and small scale model levels, there are mezzo-scale models recommended for passenger flow simulation. This kind of models consists of both generally and comprehensively characterised elements. Models can be systematised more exhaustively according to their characteristics:

- Types of modelling approach. Here we can find cellular automata and particle hopping models described in numerous publications, microscopic simulation models, queuing models, gas-kinetic models, continuum models.
- Traffic representation (individual passenger or aggregate traffic flow).
- Types of behavioural rules which can be either individual or collective unlike in the other classification method with microscopic, mesoscopic and macroscopic models.
- Variables interpretation applied in the model. Such measures as e.g. position, velocity may be presented as continuous or discrete. Some models may be classified as mixed in this field.
- Uncertainty interpretation applied in the model (deterministic with exact definition of all relations, stochastic with random variables).
- The formal assumption regarding system operation. There are models consisted of sets of equations used to more theoretical analysis and simulation models with more practical application usually containing sets of equation transformed into computer programming code as its engine.
- A designated area of a model application (e.g. in the city, between cities, between regions/nations).

From the perspective of building a simulation model to support the management of a passenger transportation system, the category of traffic representation seems especially important. Constructing models which present individual passengers behaviours require conducting a time-consuming, and expensive research in terrain which enables to collect a significant number of data. Collecting the information about a representative sample of passengers, taking into consideration diversification of the transport modes and specific features of the different parts of the transportation networks, requires the involvement of a significant number of trained people over a period of time. Moreover, the data gathered has to be stored and processed and managed. It prolongs the process of model building and increases its cost. Thus, the experts engaged in the process of model building should thoroughly analyse if there is a necessity to present passengers in micro-scale. If the data are easily accessible and already collected than the model with detailed characteristics of individual objects obviously may serve a broader set of queries, but often the needs of transportation planning authorities are satisfied with models presenting aggregated traffic flow between each pair of points in the transportation network. The macro-conceptualization of this element is a reasonable compromise between costs and utility.

The preview of the simulation models should be done in the first phase of the projects, aimed at the simulation of the process of building. It enables fixing a framework of the future work. Such a general characteristic of the model to be built enables coming to an agreement and understanding between the user and supplier. The framework of the developed tool has to be also clear to the entire project team to avoid needless overwork or the project range shifting that may result in excessive time and cost variance. Next stages of the simulation of model building require an insight into more specialist methods concerning aspects of prediction and forecasting. These approaches and their effectiveness are also inseparable from the data collection issue.

2.1. ECONOMETRIC MODELLING

Procedures and rules used as an simulation model engine can be equations, mathematical, statistical or econometric models or artificial neural networks. Algebraic models that are stochastic, including random variables (as opposed
to deterministic models which do not include random variables) are called econometric models, [13]. Econometrics is defined as: “a branch of economics in which economic theory and statistical method are used in the analysis of numerical and institutional data”, [17]. Econometric model was the first one built for the purpose of the simulation tool. It enable to verify which variables are worth to be taken to the further analysis as significantly influencing on a size of a passenger flow. Inseparable part of work with the econometric models is scenario testing. In general it can be stated that there are three main goals of developing a model:
1) Cognitive goal
2) Forecasting goal
3) Decision-making goal.

The main objective in the case of the model supporting management of passenger transportation system is to forecast the demand for public transport in the long term. The model can also be used as a computer-aided tool that supports the decision-making process of transportation demand management in the present moment to increase the comfort of passengers and benefits from the network operation. The model should be versatile and flexible. The issue of demand forecasting (including public transport demand) has been studied for many years, which has resulted in the development of numerous prediction methods and techniques, [34]. A demand forecast can be defined as the best estimate of what demand will be in the future, given a set of assumptions [31]. Forecasting techniques can be separated into two general categories – qualitative and quantitative – as shown in Table 1.

Table 1. Qualitative and quantitative forecasting methods. [27]

<table>
<thead>
<tr>
<th>Qualitative Methods</th>
<th>Time Series Methods</th>
<th>Quantitative Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judgment</td>
<td>Moving Average,</td>
<td>Regression</td>
</tr>
<tr>
<td></td>
<td>Weighted moving average</td>
<td></td>
</tr>
<tr>
<td>Historical Analogy</td>
<td>Exponential Smoothing</td>
<td>Econometric</td>
</tr>
<tr>
<td>Focus Group</td>
<td>Trend Analysis</td>
<td>Input-Output</td>
</tr>
<tr>
<td>Market Research</td>
<td>Decomposition</td>
<td>Disaggregated</td>
</tr>
<tr>
<td>Diffusion</td>
<td>Advanced Time Series methods</td>
<td>Neural Networks</td>
</tr>
<tr>
<td>Markovian</td>
<td>Box-Jenkins (ARIMA)</td>
<td></td>
</tr>
</tbody>
</table>

Qualitative prediction methods are commonly used when there is little or no data available. In general, these methods are less structured in comparison to quantitative techniques. A judgment forecasting method (eg. Delphi method) elicits management’s opinions to provide a forecast. Historical analogy obtains information from experts who have faced similar situations in the past. Another method is a focus group. It involves an objective moderator who introduces a topic to a group of respondents, and directs their discussions in a nonstructural and natural fashion. The focus group is a source of rich information that cannot be provided by a survey. However, a main disadvantage of this method is that the results obtained from a focus group cannot be generalized – the information is valid only for that homogeneous group. Diffusion models are implemented mainly to predict the fate of new products, while Markovian models to forecast consumers and buyers behaviours. Basic disadvantage of qualitative techniques is that forecasts are subjective, because they base on opinions that can be biased. Nevertheless these methods are quite popular due to their simplicity, understandability and relatively low costs.

The other group of forecasting methods are quantitative techniques. In general, these methods use historical data to predict future values, so it is often necessary to gather many data. Quantitative methods can be divided into two groups – time series methods and causal methods. Time series techniques base only on the time series data itself to build the forecasting models. So in the case of predicting future values of spare parts demand, only data regarding past demand is required. Time series methods include Moving Average (MA) and Weighted Moving Average (WMA), Exponential Smoothing (Single Exponential Smoothing, SES) and Box-Jenkins method (e.g. ARIMA). The other group of quantitative methods are causal methods.

They use a set of explanatory variables, often including time series components, that are believed to influence the predicted value, like spare parts
demand. These methods include econometrical prediction based on regression analysis and artificial neural networks (ANN), [33].

Classification of forecasting methods may be also based on the time period associated with the analyzed demand data. The models used in short-time forecasting should be simple and relatively cheap. The group of models based on the exponentially weighted average, originally suggested by Holt, has proved to be more satisfactory than any other. According to category of the forecast, certain forecasting techniques can be recommend. When considering immediate-term forecasts, applied e.g. in electricity demand forecasting, various prediction methods may be used. For short-term forecasts (e.g. demand in industry and commerce) exponentially weighted averages and derivatives are recommended. For medium-term forecasts (e.g. sales and financial forecasting) it is advisable to use regression, curve fitting or time series analysis. In case of long-term forecasts (applied in e.g. technological forecasting) methods like DELPHI or think tanks are recommended, [12]. According to [31] there are the following steps to be realized in case of applying econometric modelling in a transportation model:

- **Step 1.** System definition.
- **Step 2.** From theory-based model to data model.
- **Step 3.** Collect and refine data.
- **Step 4.** Specify functional form and stochastic structure of error term.
- **Step 5.** Misspecification tests.
- **Step 6.** Specification tests.
- **Step 7.** Evaluate the effect of uncertain exogenous variables.
- **Step 8.** Compare ex-post and ex-ante forecasts with baseline (Naive) forecasting.
- **Step 9.** Model use.

The condition of obtaining accurate forecast is a complex model. In the case of the passenger flow considered as a dependent variable, one can find in literature suggestions about at least a dozen or so of the potential factors influencing this value. The exemplary explanatory variables are:

- GDP in the analysed area,
- Average income per capita,
- The unemployment rate,
- Shuttles frequency,
- Ticket price,
- Fuel price,
- Transit time between transportation points in the given network, the
- The number of inhabitants in/around transportation points,
- Birth rate,
- The number of employed,
- The number of students,
- The number of people over 60 years old,
- The number of visiting tourists,
- Distance from the border,
- The technical velocity of the transportation modes and population density.

Data from each category have to be obtained for all points distinguished in the transportation network. It is also recommended to analyze historical sets of data to observe what trends dominated in the past and may occur significantly in the future. Similarly, as in the case of micro-scale models, building a viable econometric tool is time and cost consuming. Additionally, obtaining the historical data sometimes is just impossible if they were not collected or are aggregated in a different way than required for the need of model building. This approach implies an obligation of searching for an alternative and/or complementary methods of the effective simulation model building.

### 2.2. ARTIFICIAL NEURAL NETWORK

Artificial Neural Networks (ANN) is one of the most popular Artificial Intelligence (AI) methods, [48]. Building the ANN is the second approach that should be taken into consideration as an input to the simulation model. Artificial Neural Networks are a sophisticated modelling technique that can depict very complex functions. ANN can be defined as a highly simplified model of the structure of the biological neural network. An ANN comprises interconnected processing units. Processing unit’s general model includes a summing part followed by an output part. The summing part works as follows – it receives N input values, then weight each value, and subsequently computes a weighted sum, which is called the activation value. There are many types of Artificial Neural Networks, [39]. Basic classification groups ANN into those dedicated to classification problems (ANN(class.)) and those dedicated to regression problems (ANN(regr.)). In classification problems, the purpose of the network is to assign each case to one of some classes. Nominal output variables are used to indicate a
classification problem. The nominal values correspond to the various classes. The very often used technique is the one where there is only two-state variable. In that case, a single node corresponds to the variable, and a value of 0 is interpreted as one state, and a value of 1 as the other. In regression problems (ANN(regr.)), the objective is to estimate the value of an output variable, given the known input variables. An especially important issue in regression is output scaling and extrapolation effects. Artificial neural networks are tools applied in many areas, including sale forecasting, overhaul planning, machines diagnostics or production problems analysis, [25]. The ANN application may be grouped into three categories – applications in speech (including NETtalk, phonetic typewriter, vowel classification, recognition of consonant-vowel (CV) segments, etc.), applications in image processing (including recognition of handwritten digits, image segmentation and texture classification and segmentation), applications in decision making, [47]. The basic neuron's model is defined by the following equation, [26]:

$$y = F\left(\sum_{p=1}^{P} w_p u_p + u_0\right),$$  

(1)

Where: $u_p, p = 1, 2, ..., P$ is neuron's input; $u_0$ – bias; $w_p$ – synaptic weights; $F(\cdot)$ – activation function. An artificial neuron receives a number of inputs (from original data, or the output of other neurons). Each input comes through a connection that has a weight. Each neuron has a single threshold value. The weighted sum of the inputs is formed, and the threshold subtracted, to define the activation of the neuron. The activation signal is passed via an activation function to produce the output of the neuron. The basic types of neural network include multilayer perceptron MLP, radial basis function RBF and Kohonen network. MLP is probably the most popular network architecture in use today.

2.3. ALGORITHMIC MATHEMATICAL NOTATION

The third element that increases the probability of success in building simulation model dedicated as decision supporting tool for transportation network management is a set of formulas which describe relations, conditions and data processing procedures implemented in the model. This formal mathematical conceptualization of logics that are to be applied in the simulation system guarantees unambiguous way of communication between experts involved in the process of its development. Moreover, this method of describing dependencies and algorithms in documentation enables to easy insert modifications, even after extended period of time. Thus the mathematical record is universal and may be comprehended by another designer than those who build the simulation tool. The formulas are inherent to the econometric models and ANN, but this does not exhaust the topic of clear presentation of the complex data processing procedures in the case of integrated approach to simulation model building. There is necessity to present the elements that come from different methods cohesively with emphasize on the sequence of the data processing.

3. AN INTEGRATED APPROACH TOWARDS BUILDING A SIMULATION MODEL FOR THE PASSENGER TRANSPORTATION SYSTEM

The process of model building for transportation networks should be consistent with the classic methodology presented in [32, 36, 50]. It recommends four-stage procedure of the demand analysis in the transportation network, [36]:

Stage 1. Generating travellers flow (demand) i.e. the traffic intensity resulting from the citizens and travellers mobility.

Stage 2. The travellers dispersion in space, in particular depending on distance, time and cost of a trip.

Stage 3. The modal share (divide of the overall number of passengers depending on the transportation mode).

Stage 4. Travels division (the choice of transportation route depending on the transportation network features).

The stages are more thoroughly described in figure 1.

It is easy to notice that the requirements for the operations in each stage are rather general and leave space for the interpretation. The approach presented in this article is more complex and contains important aspect does not clearly captured in the procedure shown above. The proposed concepts integrates classical approach with the precise method of rout designation and with forecasting. The Stage III (Figure 1) was enriched with the data processing procedures explained in the following mathematical notation. Route designation in the transportation network means indicating optimal public and private route
between each pair of transportation points generating and receiving passengers flow. On this stage, the model fixes two types of routes (public, private) between each pair of points entered in the transportation network. The simulation tool, on the base of preliminary calculations made in the dedicated Excel sheet, conducts analysis of every possible combination of connections that lead from a given point \( i \) to point \( l \). The brute force algorithm searches for the optimal solution in the sequences on the base of implemented conditions and rules. The route is optimal for a passenger taking into consideration the satisfaction criterion. It is expressed in function minimizing cost and time.

Let us define a graph \( G = <N,L> \) where \( N \) is a set of nodes and \( L \) is a set of ordered pairs of nodes (lines) belonging to \( N \). Let \( n_{1}, n_{2}, ..., n_{a} (a \geq 2) \) be a sequence of distinct nodes such, that \( l_{i} = < n_{i}, n_{i+1} > \) is a line for every \( i = 1, ..., a-1 \). A sequence \( l_{1}, l_{2}, ..., l_{a-1} \) we call a route \( r_{<1,a>} \) between points \( 1 \) and \( a \). Let \( R_{<1,a>} \) be a set of all routes between points \( 1 \) and \( a \). Then for every line \( l_{i} \) we define nonnegative real functions:

\[
t_{i}(l_{i}): l_{i} \rightarrow R^{+} \cup \{0\} \rightarrow ticket\ cost\ on\ line\ l_{i} \quad (2)
\]

\[
d_{i}: l_{i} \rightarrow R^{+} \cup \{0\} \rightarrow length\ (distance)\ of\ line\ l_{i} \quad (3)
\]

We define a cost function for public transportation between points \( 1 \) and \( a \) via routes \( r_{<1,a>} \in R_{<1,a>} \) as:

\[
costR1 = \sum_{r_{<1,a>}} t_{i} \quad (4)
\]

Similarly, we define a cost function for private transportation between points \( 1 \) and \( a \) via routes \( r_{<1,a>} \in R_{<1,a>} \) as:

\[
costR2 = \sum_{r_{<1,a>}} d_{i} \theta \quad (5)
\]

The satisfaction criterion on this stage is calculated separately for the two categories of routes (public, private) as:

\[
S1 = \min_{R_{<1,a>}} (costR1 * timeR1) \quad (6)
\]

\[
S2 = \min_{R_{<1,a>}} (costR2 * timeR2) \quad (7)
\]
Where:
- cost\(R_1\) – cost of a travel in case of public routes,
- cost\(R_2\) - cost of a travel in case of private routes,
- \(\delta_{il}\) – distance between transportation point \(i\) and \(l\) given in kilometers,
- \(\theta\) – cost of the private transport per unit (kilometer), value of this parameter depends on a size of engine in a vehicle, kind of fuel and its average price. This value results from a number of connections contained in a given route and a number of transport modes used. Usually the routs serviced by one means of transport are preferred because of shorter time of travel. However, if there is not direct connection with one mean of transport, the combined solutions are allowed and indicated as optimal considering satisfaction criterion. In some cases the function of cost for the public transport is much more complex than shown above. If the public transport is subsided for certain group of passengers (e.g. students) than the formula should consider reduced fare ticket, regular ticket and the number of passengers authorized to use each kind of tickets.

Time of the travelling for the public routes (rail, bus) is calculated in the following way:

\[
time R1 = t_{DS} + t_p + t_{ch} + t_{SD}
\]  
(8)

Where:
- \(time R1\) - time of a travel in case of public routes,
- \(t_{DS}\) – time ‘door – bus stop’ – average time of getting to the bus stop/railway station and a wait time for the bus/rail departure,
- \(t_p\) – specific travel time between the initial and final bus stop/railway station,
- \(t_{ch}\) - total of the change/transfer time a bus to a train or train to a bus,
- \(t_{SD}\) – time ‘bus stop –door’ - average time of getting from the final bus stop/railway station to the travel destination.

Time of the travelling for the private route is calculated with the following formula:

\[
time R2 = t_{DC} + t_c + t_{CD}
\]  
(9)

Where:
- \(time R2\) - time of a travel in case of private routes,
- \(t_{DC}\) - time ‘door – car’ – average time of getting to the car,
- \(t_c\) - specific travel time for a car,
- \(t_{CD}\) – time ‘car-door’ - average time of parking a car and getting to the travel destination.

On the base of the classic approach recommendations, authors propose the method complemented with specific calculations and focused not only on capturing the current state but also generating prognosis for the future. The classic four-stages model of transport model building does not answer the question which transport mean is being chosen by particular number of passengers on a given route while the integrated model calculates it and displays results graphically thanks to the usage of the simulation tool. Moreover it were experimentally applied three different methods of demand calculation which is econometric model, Artificial Neural Network and Gravity Model. The complexity of the integrated approach was captured in the figure 2.

The integrated approach presented in figure 2 is built on the base recommendations presented in figure 1, but it is expanded and more accurate in the issues concerning calculations. The method was developed on the base of practical experiences gained while building a simulation model of the transportation network. The simulation tool was used to support decisions concerning transportation in Lowe Silesia region. Nevertheless the approach has universal character because it offers choice between three calculation approaches. This gives a chance to use it despite of the quality of data base gathered for the model building. The proposed integrated approach assumes a multi-step analysis. If forecasting of the passenger flow is efficient (coefficient of determination \(R^2>0.90\)) via econometric model, then a simulation model can be launched (passenger flow forecast is an input to the simulation model). If not, meaning if coefficient of determination of the econometric model is lower than 0.90, an artificial neural network is computed. And again, the efficiency of the forecasts coming from the ANN is assessed by the \(R^2\). If the result is satisfactory, a forecast of the passenger flow from the ANN can be input to the simulation model and further calculations can be activated. Otherwise, the upgraded gravity model...
is used (step III). The authors assumed that depending on the researched region or country, the availability of necessary data, and other boundary conditions, the integrated approach is the most suitable for the discussed problem. It is expected that efficiency of particular steps of the stage I can
differ depending on analyzed regions – in one case the econometric modelling can be sufficient, in other perhaps ANN development is more adequate. If none of these methods is efficient enough, a gravity model as a traditional approach may seem the most appropriate. The results of verification of the algorithms together with data processing procedures implemented in that tool are presented in the next article from this series entitled: Integrated Approach to Building Simulation Model Supporting Management of Passenger Transportation System. Part 2 – Case Study. The details concerning each stage presented briefly in figure 2 are also explained thoroughly in this publication.

4. SUMMARY

The dynamic changes in the infrastructure and demography of regions and cities enforce using dedicated tools for planning, development and management of the transportation network. To satisfy needs of citizens and attract newcomers. The authorities are obligated to dispose public budgets rationally and make their decisions on the base of confirmed facts and premises. It is especially challenging in the case of making a long time strategic plans concerning future. That is why credible modelling methods are desired and searched for. On the base of directions issued by Jaspers, there was developed the complex method of building simulation tool supporting management of the transportation network encompassing the issue of forecasting. The method comprises such an elements as econometric modelling, ANN and mathematical models. The integrated approach to building simulation model in the field of transportation considering both public and private transport modes proved to be effective both on picturing present situation, inserting changes and forecasting the future demand for the transport. The process of development of the simulation tool and testing its efficacy is described in the next article: An Approach towards Building a Simulation Model Supporting the Management of the Passenger Transportation System. Part 2 – Case Study.

REFERENCES


An Integrated Approach towards Building …


[33] M. Moon, Demand and Supply Integration: The Key to World-Class Demand Forecasting, FT Press, USA, 2013.


Date submitted: 2016-07-01
Date accepted for publishing: 2016-12-05

Joanna Gąbka
Wrocław University of Technology, Poland
joanna.gabka@pwr.edu.pl

Sławomir Susz
Wrocław University of Technology, Poland
slawomir.susz@pwr.edu.pl

Maria Rosienkiewicz
Wrocław University of Technology, Poland
maria.rosienkiewicz@pwr.edu.pl